Adopting AI to Protect Industrial Control Systems: Assessing Challenges and Opportunities from the Operators' Perspective

Clement Fung
Carnegie Mellon University

Eric Zeng
Georgetown University

Lujo Bauer
Carnegie Mellon University

Abstract

Industrial control systems (ICS) manage critical physical processes such as electric distribution and water treatment. Attackers infiltrate ICS and manipulate these critical processes, causing damage and harm. AI-based approaches can detect such attacks and raise alarms for operators, but they are not commonly used in practice and it is unclear why. In this work, we directly asked practitioners about current practices for alarms in ICS and their perspectives on adopting AI to support these practices. We conducted 18 semi-structured interviews with practitioners who work on protecting ICS, through which we identified tasks commonly performed for alarms such as raising alarms when anomalies are detected, coordinating operator response to alarms, and analyzing data to improve alarm rule sets. We found that practitioners often struggle with tasks beyond anomaly detection, such as alarm diagnosis, and we propose designing AI-based tools to support these tasks. We also identified barriers to adopting AI in ICS (e.g., limited data collection, low trust in vendor technology) and recommend ways to make AI-based tools more effective and trusted by practitioners, such as demonstrating model transparency through interactive pilot projects.

1 Introduction

Critical infrastructure (e.g., oil refineries, power grids, water treatment plants, and manufacturing) relies on industrial devices such as valves, motors, and pumps. These devices are controlled by complex, interconnected systems known as industrial control systems (ICS) [60]. As the scale of ICS grows, the potential harm caused by cyber-attacks on ICS increases [57]. Attacks on the Colonial Pipeline in 2021 and the Ukrainian power grid in 2016 highlight the potentially massive impacts of disrupting these critical industries [10, 36, 51].

To protect ICS, organizations typically use expert-defined rules to detect anomalies in ICS that might indicate imminent problems [5, 6, 11, 41]. Because rules are expensive to create and often do not detect unforeseen anomalies, researchers have proposed approaches based on artificial intelligence (AI) to detect anomalies in ICS [17, 19, 23, 35, 37, 39]. However, these systems are rarely used in practice; a recent survey found that only 10% of ICS use any form of AI on process data [15], and deployments of AI in ICS have only recently emerged [16, 54]. Suggested reasons for low adoption include cost, complexity, and data availability [29, 33, 59].

To directly investigate why AI is not commonly used in ICS and to explore new opportunities for adopting AI to protect ICS, we conducted 18 semi-structured interviews with practitioners who work on monitoring, operating, and securing ICS in various industries and roles. Based on these interviews, we identify tasks commonly performed for alarms in ICS as part of an alarm workflow. Alarm workflows often involve defining rules to detect anomalies, reading real-time data from the ICS and raising alarms, responding to alarms, and modifying alarm rulesets for efficiency and safety. We answer the following research questions:

- **RQ1:** What types of data and systems are used for alarms in ICS, and how suitable are they for AI?
- **RQ2:** What human tasks are performed with alarms in ICS, and how can AI support them?
- **RQ3:** In organizations that operate ICS, what logistical and cultural factors hinder AI adoption?

In answering these research questions, one particular challenge is that practitioners working with ICS typically do not have experience with AI, which limits their ability to provide details on how AI could be used to protect ICS. Thus, in our interviews, we first performed a needs assessment of current

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.
USENIX Symposium on Usable Privacy and Security (SOUPS) 2025.
August 10–12, 2025, Seattle, WA, United States.

practices for alarms in ICS. We asked practitioners how they design, use, and maintain systems to raise alarms; how they coordinate alarm response; and what challenges they experience with alarms. Since practitioners in ICS typically do not use AI, we next asked about what benefits and barriers they perceive to adopting AI in ICS.

Although most prior work that proposes AI for ICS security focuses on centralized AI models for detecting anomalies [23, 33], our findings suggest that other use cases for AI are likely to be more promising in practice. We found (**RQ1**) that data and systems for raising alarms are often not centralized, but historical data from alarms is. We also found (**RQ2**) that practitioners often struggle with tasks beyond detecting anomalies, such as diagnosing alarms and managing alarm rulesets. We therefore propose creating tools for diagnosing and managing alarms on centralized, historical data. Furthermore, we found (**RQ3**) important cultural barriers to deploying and using tools in ICS, such as general skepticism towards adopting new technology. We therefore recommend ways to navigate these barriers; for example, given the importance of trust in ICS, we recommend that tool designers build trust with practitioners by interactively demonstrating how AI-based tools make predictions.

This paper is structured as follows: we review background in Sec. 2 and describe our study methodology in Sec. 3; we convey our findings for current practices for alarms in ICS in Sec. 4 and practitioners' perceptions of using AI in ICS in Sec. 5; finally, we answer our research questions and provide recommendations for adopting AI to protect ICS in Sec. 6.

Connecting our work to cybersecurity: Our primary interest in this investigation is learning how to help detect attacks on processes controlled by ICS. At the process level, detecting these attacks is intertwined with detecting non-malicious anomalies. Hence, our investigation by necessity examines participants' perspectives on the detection of all anomalies, and not just those caused by attacks.

2 Background and related work

In this section, we provide background on ICS (Sec. 2.1) and describe related prior work at the intersections of ICS, AI for cybersecurity, and human factors (Sec. 2.2).

2.1 Industrial control systems

An ICS monitors and controls a physical, industrial process. ICS use programmable logic controllers (PLCs) to read process information from sensors (e.g., temperature, pressure, or flow sensors), execute controller logic, and send commands to actuators (e.g., valves, pumps, or motors). To coordinate and communicate between multiple PLCs, systems such as supervisory control and data acquisition systems (SCADA), distributed control systems (DCS), or human-machine interfaces (HMIs) are often used. ICS devices are commonly organized according to the Purdue model [31], which defines a hierarchy based on logical proximity to the physical process. Fig. 1 shows a typical categorization of ICS devices in the Purdue model, including sensors and actuators (level 0); PLCs (level 1); SCADA, DCS, and HMI (level 2); database and analysis functions (level 3); and business-level functions such as email (level 4 and beyond). Systems at levels 0–3 are commonly referred to as operational technology (OT), and systems at levels 4 and higher are commonly referred to as information technology (IT) [44].

Given their critical nature, adversaries have strong incentives to attack ICS for harm or profit. For example, the Colonial Pipeline attack in the United States disrupted oil production for several days and cost over 4 million US dollars in ransom [10], attacks on the Ukrainian power grid caused over 200,000 people to lose power for several hours [36,51], and an attack on a German Steel Mill caused equipment damage [70].

2.2 Related work

In this section, we provide context for our study by describing related work that proposes AI-based anomaly detection, related work that studies security practitioners' perceptions of AI-based tools, and related work that studies practitioners who work with ICS.

ICS anomaly detection. To protect ICS from harm, researchers have proposed process-level ICS anomaly detection systems based on rules [1, 22], physics-based equations [26, 66], statistical modeling [4, 28, 62], and deep learning [23, 35, 37]. Although AI-based anomaly detection is commonly proposed in research [17, 19, 23, 33, 35, 37, 39] and some deployments of AI in ICS have been reported in practice [16, 54], no prior work has broadly studied ICS operators' perspectives of these AI-based approaches. In this work, we investigate operators' perspectives on how anomaly-detection systems are currently used and opportunities for AI to improve current practice.

Alarm workflows in security operations centers. Security operations centers (SOCs) are organizational units responsible for securing an entire organization, focusing on information technology (IT) networks and systems. A variety of prior work studies the challenges that SOC operators face in their day-to-day roles: alarm response [2], alarm ruleset management [63], and organizational challenges [34]. Alarm workflow tasks in SOCs are similar to those performed in ICS, but they interact with fundamentally different technologies. SOCs and IT systems operate at high levels of the Purdue model (i.e., level 4) and do not directly interact with operational technology (OT), such as PLCs or SCADA. IT and OT professionals exhibit different cultural beliefs about ICS security [25], and our work focuses on alarm workflow tasks for OT systems, performed by OT professionals.

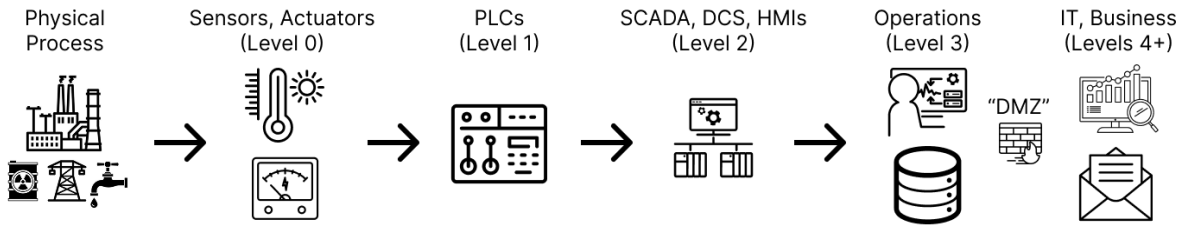

Figure 1: We show how ICS devices are categorized in the Purdue model of ICS [31]. Generally, devices further from the industrial process are categorized into higher levels. For example, sensors that directly read process information are at level 0, PLCs that read data from sensors are at level 1, a DCS that manages multiple PLCs is at level 2, and a historian database of process information is at level 3. ICS commonly use firewalls to separate levels 0–3 from higher-level IT systems, also known as the "demilitarized zone" (DMZ).

Security practitioners perceptions' of AI-based tools. Prior work has studied the perceptions of IT security practitioners, either focusing on their perceptions of a specific AI-based tool [46, 48] or by studying their perceptions of AI in general [42]. These works identify promising uses for AI-based explanations to help practitioners understand security events, but also identify concerns with accuracy, trust, and usability. In organizations that manage ICS, we found that dedicated roles for OT security are uncommon, as described in Sec. 4.4. We therefore focus on making recommendations for AI-based tools aimed at OT practitioners in ICS.

Working with ICS. Prior work has also studied the perspectives of practitioners who work in ICS operations [9, 56] and ICS security [25, 55], identifying various technical, cultural, and organizational challenges. ICS can suffer from a lack of standardization [9, 55, 56], computational constraints [9], and tensions between IT and OT professionals [25]. Furthermore, these works found that practitioners who work with ICS often mistrust vendor tools [56] and disagree on the significance of cybersecurity threats [55]. In our work, we observed similar challenges to those reported in prior work (Sec. 4.4), but we focus specifically on how these challenges affect the adoption of AI for alarm workflows in ICS.

3 Participants and methodology

We interviewed 18 practitioners who work with ICS in multiple industries and roles. In this section, we describe: how we recruited participants (Sec. 3.1), how we conducted semi-structured interviews (Sec. 3.2), our methodology for analyzing interview responses (Sec. 3.3), our consideration of research ethics (Sec. 3.4), and study limitations (Sec. 3.5).

3.1 Participant recruitment and demographics

Our target population is practitioners who work on safeguarding and securing ICS by performing alarm workflow tasks or by managing or supporting alarm workflows. To recruit participants from this population, we used purposive sampling. We directly contacted individuals in our professional networks; we advertised on ICS security mailing lists; we emailed utility providers with public contact information; we posted flyers on LinkedIn and X (formerly Twitter); and we sent direct messages on LinkedIn to people whose roles matched ICS-related keywords such as "SCADA" or "Control System."

In our initial recruitment text, we used the terms "anomaly detection" and "security," and we failed to recruit participants; multiple organizations responded that they did not perform anomaly detection or did not have any security-relevant topics to discuss. Given the sensitive nature of cyber-attacks on critical infrastructure, we believe that participants were unwilling to discuss these topics or believed that they were not relevant to them. We then updated our recruitment text to state that we were interested in "monitoring tools" and "alarm response." We were then able to successfully recruit study participants and discovered that, in fact, they do use systems to detect anomalies and acknowledge that cybersecurity concerns can impact alarm response. This experience serves as a useful lesson that, when interacting with practitioners who work with ICS, using appropriate terminology is important.

Potential study participants then filled out a screening survey, which asked about their industry, day-to-day tasks, and experience with ICS, cybersecurity, and AI. We screened participants for those who demonstrated experience with operating, managing, or developing tools for alarm workflows. We also limited study participants to those located in the USA, although some participants reported on prior experiences from working in other countries.

Table 1 provides an overview of study participants' demographic information. Participants worked for two different types of organizations: *plant owners*, organizations that operate an ICS, and *vendors*, organizations that support alarm workflows for multiple ICS. Of the participants who worked for plant owners, six participants worked for local municipalities across five different US states.

ID	Industry	Role	# Years Exp.		
			OT	Sec.	AI
P1	Electric (solar)	Manager	10	0	10
P2	Oil & Gas	Engineer	1	0	0
P3	Electric (grid)	Engineer	12	4	6
P4	Water	Manager	10	0	0
P5	Water	Engineer	2	0	0
P6	Water	Manager	15	10	0
P7	Oil & Gas	Manager	18	0	0
V8	Electric (gen.)	Consultant	50	24	0
P9	Manufacturing	Engineer	13	0	0
V10	Electric (gen.)	Engineer	20	15	4
P11	Electric (grid)	Manager	10	10	0
V12	HVAC	Engineer	5	2	0
P13	Oil & Gas	Engineer	4	0	0
P14	Manufacturing	Engineer	7	0	0
V15	Electric (grid)	Engineer	25	10	0
V16	Oil & Gas	Consultant	8	13	0
P17	Oil & Gas	Engineer	16	14	4
P18	Water	Manager	35	10	0

Table 1: Demographic information for the 18 participants in our study: their industry; their role; and their years of experience with operational technology (OT), cybersecurity (Sec.), and AI. 13 participants primarily worked for a plant owner that operates one ICS (marked "P"), while five participants primarily worked for a vendor or as a consultant that supports multiple ICS (marked "V").

Similar to challenges reported in related work [2, 63], we found it difficult to recruit operators who worked as the first point of contact in alarm response for an ICS. Our interviews revealed that ICS are often monitored 24/7 by operators who are often overworked. Thus, these operators likely could not provide the time to participate in an interview for research purposes. Although we could not recruit operators who worked as the first point of contact in alarm response at the time of recruitment, the participants in our study manage these operators, perform secondary alarm response tasks, or worked as operators in the past. Thus, participants were able to discuss operator perspectives through second-hand experience and prior first-hand experience.

As described in Sec. 3.3, we iteratively performed qualitative analysis after establishing an initial list of codes. We determined that recruitment was complete once we observed that no new qualitative codes directly pertaining to our research questions emerged (inductive thematic saturation [52]).

3.2 Interview methodology

We conducted 60-minute, semi-structured interviews over Zoom. Participants filled out a consent form before starting the interviews. We recorded interviews with participants' consent or took notes if participants did not consent to recording.

We divided our interview into four sections. In part I, we asked the participant about their professional background and day-to-day responsibilities. In part II, we asked about current practices for alarms in ICS: how data is collected and alarms are raised, how alarm response is performed, and how these processes are managed. In part III, we asked about vendor tools and how they are adopted in ICS. Finally, in part IV, we asked participants about their perceptions of using AI in ICS. When interviewing participants who worked for vendors, instead of asking about the practices of a single ICS, we asked about common practices and trends observed from working with clients. To ensure question clarity, we performed a pilot test of our interview script with two researchers (who are not directly involved with this work) with experience in human-subjects research in ICS contexts. Our interview script can be found in Appendix A.

3.3 Analysis methodology

To prepare our data for analysis, we transcribed interview recordings using an automatic transcription service. We edited all transcripts and notes for correctness and anonymity by removing specifically identifying information related to people, places, and companies before deleting the original recordings.

To analyze our interview data, we used inductive thematic analysis [12]. After completing the first 16 interviews, two researchers iteratively and independently reviewed each transcript, creating a list of initial codes that captured concepts relevant to our research questions. Once all interviews were complete, the two researchers met to merge codes and group related codes into themes, producing our initial codebook. We then refined the codebook using an iterative, consensus-based approach to ensure that the two researchers shared a conceptual understanding of the codes and could apply the codes consistently. The two researchers independently coded two transcripts using the initial codebook, met to identify disagreements and update code definitions, and resolved all discrepancies in the codes. Using the refined codebook, both researchers independently coded four more transcripts, met to discuss codes, and found no substantial disagreements on the definitions or usage of the refined codes. After reaching consensus on the codebook and its application, one researcher subsequently coded the remaining interviews. To ensure consistent application of codes, another researcher reviewed these codes for correctness. Our final codebook and code counts are in Appendix B.

Following suggested practices in qualitative HCI research [40, 47], we do not compute inter-rater reliability metrics since the goal of our study is to identify emergent themes rather than to quantify the frequency of topics. We ensured consistency by refining the codebook when disagreements were found in double-coded interviews and reviewing each single-coded interview. Furthermore, we do not report the exact counts of participants when discussing results since our

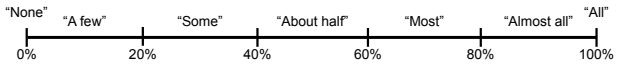

Figure 2: When reporting the proportion of participants in results, we use qualitative terms instead of raw counts or percentages. We convert percentages to terms based on the scale shown, using a mapping similar to that of prior works [21,27].

primary findings are qualitative and we gathered perspectives from a diverse but not necessarily representative sample of practitioners. We instead follow a common methodology from prior work and use qualitative terms to illustrate the prevalence of themes [21,27], by mapping percentages to terms as shown in Fig. 2.

3.4 Ethics

Our study was approved by our institution's Internal Review Board (IRB). Participants were compensated $60 USD for completing interviews, a similar rate to prior work that interviews domain experts [32,38,42]. Following best practices, we minimize participant harm by obtaining informed consent, anonymizing transcripts, and asking participants not to share confidential information about their organization or role [7]. We follow these principles to protect participants' individual privacy, to protect participants from potential repercussions from their employer, and to protect their employers from increased risk by disclosing sensitive information about cybersecurity practices.

We also weighed the benefits of publishing this work, which reveals security practices in ICS, against the risks of releasing more information to adversaries. We concluded that the risks were minimal since ICS are already commonly attacked [57], and industry surveys already disclose that AI-based tools are not commonly used in ICS [15].

3.5 Limitations

The responses of participants we interviewed may not fully represent the perspectives of current, first-response ICS operators, since they do not currently serve as the first response to ICS alarms. Some participants in senior roles had not worked as operators for several years, and their responses may not fully represent all operators due to organizational communication barriers and changes in the industry. Furthermore, all study participants were based in the USA, which may limit the applicability of our findings to other countries.

We describe participants' suggestions for AI adoption in Sec. 5. Most participants lacked experience with AI, and so potential misconceptions about the requirements and capabilities of AI may affect the feasibility of their suggestions.

4 Results: Current practices for alarms in ICS

In this section, we report our findings for how ICS operators use systems for alarms and perform alarm workflow tasks. Since AI is not commonly used in ICS and most practitioners who work with ICS lack experience with AI [15] (including the participants of our study, shown in Table 1), we use an indirect approach to investigate our research questions by first asking about current practices for alarms in ICS. As a pre-requisite for answering **RQ1**, we ask participants about systems that read data from an ICS and raise alarms (Sec. 4.1). To support our investigation of **RQ2**, we ask participants about human tasks performed for alarms (Sec. 4.2) and challenges with performing these tasks (Sec. 4.3). Finally, for our analysis of **RQ3**, we ask participants about ICS-specific factors that affect alarm workflows (Sec. 4.4) and adopting vendor tools in alarm workflows (Sec. 4.5). Although the individual processes used for each ICS vary, we identify common processes for alarms across ICS, and we show a categorization of these processes in Fig. 3.

4.1 Systems for raising alarms

Anomaly-detection systems, whether AI-based or rule-based, rely on real-time data from the ICS, so understanding how this data is collected is critical to understanding how AI can be used for alarms in ICS.

What devices and systems are used? Referring to the devices described in Sec. 2.1, almost all participants who worked for plant owners use PLCs. Almost all participants also use a level 2 system to coordinate multiple PLCs—some participants use a DCS and some participants use SCADA. About half of participants report using one or more control rooms, which are centralized locations for operators to monitor ICS and delegate alarm response. In contrast, some participants reported that their organization does not use control rooms; operators instead interact with PLCs through co-located HMIs, which are distributed across the industrial process. Finally, some participants report that their organization uses a data historian, which stores process and alarm data in a central database for post-hoc analysis. We describe these post-hoc analysis tasks in Sec. 4.2.

Which organization manages these devices? Although plant owners use various devices for monitoring and controlling industrial processes, they do not necessarily program or manage these devices. Some participants work for plant owners who rely on vendors to manage their devices; a few participants reported that this was common in their industry. In contrast, some participants work for plant owners that employ their own staff to program and manage their devices. We describe how vendors affect alarm workflows in Sec. 4.5.

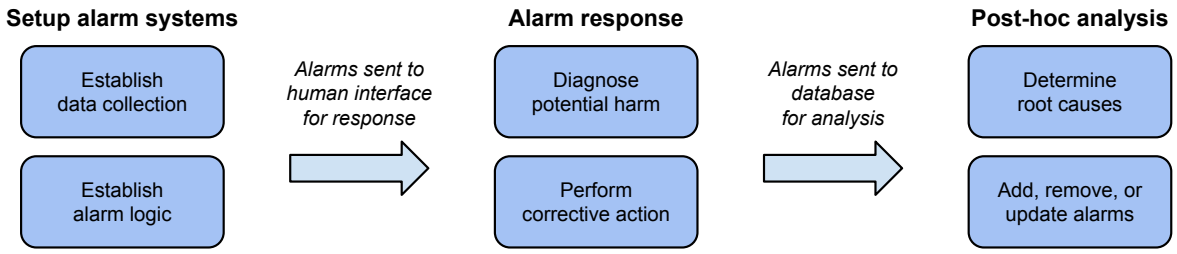

Figure 3: We found that alarms are managed through a set of processes in an alarm workflow. We show the different tasks in alarm workflows, categorized into three stages: (i) setting up systems for alarms, (ii) responding to alarms, and (iii) analyzing alarms post-hoc to determine root-causes and update alarm conditions.

> *A lot of them actually contract out their PLC, SCADA, networking, some of the more high level stuff. Actually, almost every city that I know does that. –P6*

What behaviors are alarms used for? All participants reported using alarms to detect anomalies in process values, although the reasons for detection varied. Most participants mentioned safety: process alarms ensure the safety of the physical process or the process equipment. Some participants mentioned non-critical reasons, such as to ensure adequate production or to ensure that regulations are being met. Some participants also reported using alarms that were not related to process values; these include alarms to detect component failures (e.g., an unresponsive PLC), cybersecurity events, or physical security events. Finally, some participants reported special types of alarms, such as informational alarms, maintenance alarms, and alarms written for specific, prior incidents.

How are alarms defined? Most participants reported that alarms were defined by rules; alarms were raised if a process value exceeded an upper or lower limit. About half of participants additionally reported that more complex logic was used, such as using rates of change or combinations of rules.

> *We have combined conditions to generate a new alarm or suppress some alarms. For example, if we trip something, we don't need to see alarms from every downstream device. –P17*

Where is alarm logic implemented? Since various types of devices are used in alarm systems, the placement of alarm logic also varies. Most participants reported that alarm logic was written in PLCs; these alarms only evaluated conditions based on data available to the PLC. Some participants reported that alarm logic was instead written in DCS or SCADA; these alarms were often more complicated and required data from multiple PLCs. Finally, most participants also reported that alarms were written directly into level 0 devices (e.g., sensors), referred to as "safety systems." Safety systems can perform commands (e.g., shutting off a valve) without human involvement or inter-device communication and often send alarms to higher-level systems for diagnosis.

Where are alarms displayed? Although alarm logic is written into devices at levels 0–2, alarms are not necessarily displayed in these devices. Some participants reported that alarms from PLCs were forwarded to DCS, SCADA, or HMI. Devices in levels 0–1, such as safety systems and PLCs, often lack an operator interface, so alarms from these devices are forwarded to human operators in a level 2 system.

> *The SCADA is pulling from the PLC and if there's an alarm, it's going to display that on the SCADA itself. –P14*

However, a few participants reported that alarms were not always forwarded. In some cases, alarms could be raised without visibility to a level 2 system.

> *Some of those alarms will not go to SCADA, at least not directly. [. . .] If a relay causes a breaker to open, the relay knows why it opened, SCADA would see the breaker open, but if you were looking at your SCADA logs, you would never get any indication as to why that breaker opened. –V15*

Takeaways for AI. Participants reported using a variety of devices for raising alarms, which can range in data availability and computational power. Additionally, some plant owners rely on vendors to manage these devices. These differences make it unclear who would manage an AI-based tool for alarms, and where it should be deployed in an ICS. We also found that alarm conditions use logic and implementations that may differ from AI-based anomaly detection. For example, alarms use various data modalities (e.g., network and process data) and custom logic that may not correspond to learnable patterns in a dataset.

4.2 Human tasks in alarm workflows

We asked participants about human tasks performed in alarm workflows, such as responding to alarms and managing alarm rulesets. We investigated if and how humans performing these tasks could be supported by AI.

Who responds to an alarm? Most participants reported that an on-site operator is the first to respond to an alarm. For

a majority of alarms, operators are able to respond appropriately, either by performing the required remediation action or by acknowledging the alarm as a non-issue. If the proper response could not be determined or performed, operators would then escalate to higher levels of authority for help. A few participants who worked for plant owners also reported contractual agreements with vendors for alarm response. A few participants reported that they served as the second or third point of alarm response for an ICS.

> *The alarm does not go away or it gets worse, then you escalate up to the next line. I'll definitely get involved in troubleshooting and trying to figure out stuff like that. –P13*

How is the response to an alarm determined? Most participants reported using structured protocols to remove ambiguity in alarm response. About half of participants reported that alarm response was dictated by pre-defined categories for alarms.

> *The operator knows if an alarm comes in color red, you have to address that right away. If it comes in this color, you just call this person. If it comes in this color, you don't even have to do anything. –V15*

However, structured protocols do not cover all cases of alarm response. Most participants reported that alarm diagnosis sometimes relies on operator expertise—operators diagnose alarms by correlating them with auxiliary data in an intuitive, unstructured process. Despite using structured alarm response protocols and auxiliary data sources, less-experienced operators can struggle with alarm diagnosis.

> *I think that's probably our greatest challenge: training the staff that's still fairly new and still learning the processes what the appropriate level of response is. –P18*

How are alarms analyzed post-hoc? In cases where the real-time alarm diagnosis and response was incorrect, organizations analyze historian data for root-cause analysis. About half of participants reported that their organizations have specific teams or roles for asynchronous, post-hoc alarm analysis.

> *Another team is looking through our alarm history and identifying where we have ongoing issues or where we didn't respond to something the way we should have. –P7*

A few participants reported analyzing alarms post-hoc for alarm management. Participants mentioned the ANSI-ISA 18-2 standard on alarm management [30] and discussed alarm management tasks such as reviewing and updating alarm conditions to reduce operator fatigue. Organizations perform alarm management by reviewing historical alarm data through regular, cross-functional meetings to ensure that existing alarms are effective. If needed, alarms are added, removed, updated, or re-categorized.

> *We have alarm management expectations. Once a week, as an engineering team, we meet and review all of the alarms that came in over the last week, and try and figure out, was this a good, useful, real alarm? [. . .] And you can make changes to the alarm set points or things like that. –P13*

Some participants did not explicitly mention "alarm management" but reported other processes for managing alarm rulesets. Some organizations regularly test alarms, some organizations use tools to analyze alarm data, and some organizations allow operators to update alarms themselves.

> *We use an alarm analysis tool. [. . .] It's doing SQL queries to find repeat offenders or numbers per hour. –P7*

Finally, a few participants explicitly suggested that, since their historian data was centralized and labeled, an AI-based tool could be trained to help with alarm analysis.

> *So we now have 1000s upon 1000s of examples of: the data was doing this at the time, it led to this root cause analysis, and it led to this action. And I think that's something that we can begin to look at applying deep learning algorithms to, because we have the necessary data to start training that. –P1*

Takeaways for AI. Although most alarm response is performed through structured protocols, we found that some tasks for alarms rely on intuition and expertise, such as alarm diagnosis and alarm management. These tasks involve post-hoc analysis of alarm data and already include automated processes and tools, which suggests openness towards using AI for these tasks.

4.3 Challenges with alarms

In this section, we describe common challenges reported by participants when working with alarms. Participants reported challenges across alarm workflows, from correctly diagnosing alarms to deciding if alarm rulesets were effective.

Nuisance alarms. Most participants discussed "nuisance" alarms, which do not correspond to genuine harm, do not convey useful information, and are not actionable. Operators are burdened with recognizing and acknowledging nuisance alarms to remove them from user interfaces, and some participants reported that it can be difficult to distinguish nuisance alarms from genuine alarms. When asked about the frequency of alarms at the ICS they worked with, participant responses ranged from two to 250 alarms per hour. A few participants mentioned a growing awareness about alarm fatigue and its potential to increase the likelihood of operator errors.

> *Processes may have some variances that are going to cause nuisance alarms. Eventually, they run into, "Oh, I've ignored too much. Now I've got myself into a hole." So we try to be very judicious about what we what we alarm about. –V10*

Some participants reported using strategies to mitigate nuisance alarms. These include alarm management (described in Sec. 4.2), defining guidance for alarms (e.g., "alarm philosophy"), using tools to suppress nuisance alarms, and using tools to identify nuisance alarms. One participant described their experience reviewing and updating alarm rulesets at their organization, reporting that it required significant amounts of time and labor.

> *We looked at every single alarm that we have, and then went through and wrote troubleshooting guidance, and then challenged if you need the alarm, and then what the alarm point should be. And that was a significant year and a half of, at least 10 hours a week. –P13*

Challenges in alarm diagnosis. Even if an alarm is determined to be genuine, alarm diagnosis can still be challenging. Some participants reported limited access to data as a challenge for alarm diagnosis. In some cases, due to missing coverage in monitoring or logging, operators would need to physically travel to properly diagnose an alarm, limiting their ability to respond quickly.

> *It could be 10 minutes to an hour or two, in some parts of the country, of drive time before you even get eyes to see what's going on. –P11*

A few participants reported that too much information could overwhelm operators and hinder diagnosis. Providing data to operators requires a balance between sending sufficient diagnostic information and avoiding information overload.

> *We weren't rationalizing what we were bringing in, trying to bring everything back that we could. And so getting any value, really, out of those alarms was difficult. –P7*

Takeaways for AI. Participants reported challenges with alarm workflow tasks, particularly for alarm management and alarm diagnosis. We suggest that AI could help mitigate these challenges and improve these tasks, and we describe participant perspectives on using AI for alarm workflow tasks in Sec. 5.2.

4.4 Factors that affect alarm workflows

In this section, we describe factors specific to ICS that affect how alarm workflows are designed and executed. We identify opportunities for improvement and pitfalls that should be avoided when using AI for alarm workflows in ICS.

Limited resources in ICS. Plant owners often have small OT and cybersecurity budgets [49], introducing constraints on technology and personnel. Most participants discussed how these constraints made alarm workflow tasks more difficult.

About half of participants who worked for a plant owner reported personnel constraints, including understaffed teams and limited technical skills. These constraints limited plant owners' abilities to improve their processes for alarms, as they were occupied with critical operations and alarm response.

> *I've got one assistant now and I feel like I could keep three assistants busy. There's this triage of things that would be valuable to do versus things that are urgent. –P4*

Participants also reported challenges with managing or using technology in ICS. Some participants reported challenges with managing OT networks, which caused problems with data visibility and trust. A few participants reported concerns with device capabilities, in terms of both computation and networking. A few participants also reported challenges with updating and replacing ICS devices. Given these technical constraints, some participants expressed doubt that adopting advanced tools for alarms would help.

> *They wrote a little report on it. Here's some potential alternatives. [...] But the alternatives may not work any better because the problem may really be that our network is unstable. –P4*

Perceptions of safety in IT vs OT. About half of participants reported tensions between informational technology (IT) and operational technology (OT) professionals; such tensions have been studied in prior work [25, 67, 69]. We found that these tensions can add friction to alarm workflow tasks when IT professionals interact with OT systems.

> *People that came up from the IT side, their mental model of digital systems, it's not working, I'll just reset it. I can't do that if I'm running a power grid. [...] Because IT capabilities are continuing to get pushed closer to physical systems, people are coming from the IT side. That's where I think a majority of the mental model change needs to happen. –P11*

We found that safety, rather than security, was the focus for most participants. About half of participants acknowledged that ICS security was relevant to their work, since attacks on ICS could affect systems used for alarms. However, some participants reported that their organization does not use OT security tools, instead reporting that they believed security was the responsibility of IT.

> *A lot of the industrial control equipment, your PLCs and stuff, are pretty vulnerable. If somebody can get to them over a network, we feel like we're already essentially screwed. So the emphasis is more on the IT side and keeping those things off the network. –P4*

Of the participants who work for vendors, almost all reported that concerns about security emerged in their discussions with plant owners, which affected their practices for monitoring and connectivity.

> *Some people will not connect [relays] to SCADA because there's a worry if your SCADA system had a problem, mainly a cyber problem. [...] By isolating it completely, you've got another level of confidence. –V15*

Government regulations. In some industries, government regulations mandate which monitoring, alarms, and security practices must exist. Of the four participants who worked for plant owners in the water industry, half reported that they wrote alarms for certain process values because they were required by regulation. Most participants who worked in the electric industry reported on how NERC CIP (North American Electric Reliability Critical Infrastructure Protection) standards [61] impact their systems for alarms. In some cases, regulations would cause plant owners to consider the trade-off between increased monitoring and the additional cost of required compliance.

> *Does the device have connectivity? If it does have connectivity then you've got to follow these extra rules. But if you can say the device has no connectivity, then you don't have to answer those next questions. –V15*

A few participants also reported on differences in regulation across industries. One participant commented specifically on how regulation could affect AI adoption, stating that NERC CIP would serve as a barrier for adopting AI in the electric industry. Developers of AI-based tools will therefore need to meet NERC CIP regulations for adoption in the electric industry.

> *If it's like oil and gas, where it's a well-known, well-documented process, they're more likely to embrace AI type stuff. If it is a power-gen aeroderivative turbine, a lot of times it's going to be very human centric, because they cannot document the AI-ness for NERC CIP. –V10*

Takeaways for AI. Several factors can limit how alarm systems are used and how alarm workflow tasks are performed: difficulty updating devices, mismanaged OT networks, understaffed and undertrained teams, cultural friction between IT and OT professionals, and restrictive government regulations. Tool designers must successfully navigate these barriers for effective adoption in ICS alarm workflows.

4.5 Adopting vendor tools in ICS

In Sec. 4.1, we reported that some plant owners have contracts with vendors for monitoring and alarm response. This suggests that vendors could potentially be the driving factor towards adopting AI for alarm workflows in ICS. Since most existing tools are not based on AI, we asked participants about how vendor tools in general are evaluated and adopted into ICS. We discuss participant responses when asked specifically about adopting AI in Sec. 5.

What barriers hinder adoption? Almost all participants reported that adopting vendor tools in ICS is difficult. Participants reported concerns with vendor tools such as high cost, requirements for skilled personnel, lack of customization, and a lack of trust.

> *You need personnel, you need people trained on new technology, if you want to put in new technology. And that was a problem with certain brands. –P5*

Some participants reported that plant owners prefer to keep their systems homogeneous under a single vendor; many vendors provide solutions across the ICS stack for control logic, alarms, and OT networks.

> *There's better stuff out there that we could be using, but our investment in <Brand A> versus what it would cost and the amount of work it would take me to switch over to <Brand B> is pretty unfeasible. –P6*

What values are important for adoption? Participants reported that quantitative criteria were not frequently used to evaluate vendor tools. Some participants reported that metrics like accuracy or F1-scores, commonly used in ICS anomaly detection research [23], were not meaningful to them. Instead, almost all participants reported that tools were evaluated qualitatively. Values such as brand reputation, positive discussions with vendors, and positive recommendations were reported as most important. A few participants mentioned vendors that provide AI-based tools for ICS, but reported that they were mostly not yet trusted by the industry.

> *You could do all of this testing to say, what's the percentage of identified anomalies versus unidentified anomalies? There's things like that. Yeah, not at my level. –P8*

Who decides to adopt new technology? Finally, we found that who made tool acquisition decisions varied. Some participants reported that a cross-functional team decided if a vendor tool was adopted, whereas some participants reported that practitioners who work with ICS, including the potential end-users, were often excluded from these decisions.

> *Some of that decision making, sadly it's going to be some really slick talking salesman talking to an engineer that will never work with that SCADA system. And then the people that use that SCADA system aren't going to have a say in what they're using. That's just the way it is. –P6*

Takeaways for AI. Tool adoption in ICS is heavily based on trust and reputation. Detection-based metrics, which are used to compare AI models in research, are not used to motivate adoption. Tool designers and vendors should therefore develop new metrics that better convey trust and build their reputation with ICS insiders.

5 Results: Perceptions of AI

In the final part of our interview, we asked participants directly about their perceptions of using AI in ICS. We allowed participants to respond based on their own conceptual model of AI, but since most study participants had no experience with AI

(shown in Table 1), we acknowledge that these suggestions may not necessarily be practical.

We describe participants' conceptual models of AI (Sec. 5.1), perceived benefits of adopting AI (Sec. 5.2), and perceived barriers to adopting AI (Sec. 5.3). Combined with our findings of current practices and challenges (Sec. 4), these responses reveal opportunities and challenges for using AI to support alarm workflows in ICS and guide our recommendations in Sec. 6.

5.1 Conceptual models of AI

Given the lack of participant expertise in AI, we first establish and describe participants' mental models of AI. In this study, we define AI to be any technology that uses historical data to learn patterns and makes predictions on new data, such as neural networks [17, 23, 35], large language models (LLMs) [58], SVMs [3, 53], and confidence intervals [18]. We found that all participants demonstrated some understanding of what AI was and its capabilities. Most participants mentioned a specific model, such as LLMs, neural networks, or linear regression models. The remaining participants did not mention a specific model, but correctly referred to AI as technology that learns and makes predictions from data.

5.2 Perceived benefits of adopting AI in ICS

Some alarm workflow tasks involve processing large amounts of data and seeing complex patterns. Based on their understanding of AI and its capabilities, most participants thought AI was well suited for such tasks.

Making alarm workflow tasks more efficient. About half of participants believed that using AI would save operators' time. In Sec. 4.4, we reported that ICS operations teams are often understaffed and fatigued, and AI could help reduce this burden. About half of participants believed that AI could outperform humans at some tasks, such as seeing complex relationships in data or avoiding distractions.

> *If there was some kind of machine learning, it might notice things like, these alarms happen when you're running that motor over there which you wouldn't think is related. –P4*

A few participants who worked for plant owners reported trying LLMs for alarm diagnosis, outside of their organization's established alarm workflows. These participants tried using LLMs to diagnose previously resolved incidents. Participants reported positive experiences with LLMs, which were able to correctly diagnose the incidents and strengthened their belief that AI could provide value in similar circumstances.

> *In the prompt, I put the question: we know that a certain equipment was tripped, so we asked to find why it's tripped. . . . We spent like a whole bunch of hours, but this model for 10 seconds, and having 30% of information we had, gave us the cause of the trip. –P17*

5.3 Perceived barriers to adopting AI in ICS

In Sec. 4.4, we described barriers to adopting vendor technology in ICS. Echoing these findings and based on their understanding of what AI is and its requirements, participants reported their perceived barriers to adopting AI in ICS.

Limited compute and data availability. Participants reported concern that using an AI-based tool would require data and computational infrastructure beyond their organization's capabilities. About half of participants reported that data availability was a barrier to adopting AI because their data quality is too poor, their data is not sufficiently labeled, or that there would be IP issues with AI-based tools accessing their data.

> *The mistake would be: Hey, we have this incredible system that can detect all these problems. [. . .] But no site manager wants to set them up with the massive data requirements to make that product run. –P1*

Limited people with AI expertise. Some participants reported concerns with finding and hiring the specialized staff required to use AI-based tools. On the other hand, a few participants suggested that adopting AI could help with staffing issues, suggesting that using AI could attract a younger, more skilled workforce to ICS.

> *You need the person who knows how the process is controlled, how alarms are generated, and how an LLM works, which is not easy to find. –P17*

Low trust in AI. Most participants reported that, since ICS are so critical, they needed tools that were trustworthy. About half of participants reported concerns with AI's lack of transparency or tendency to make errors. These concerns made it difficult to convince plant owners to adopt AI-based tools.

> *The customer, he doesn't trust [AI] in control at all. [. . .] If you give them a method which has some problem with robustness, it can cost him millions if he got to shut down activity. –V12*

When asked about ways to improve AI in general, some participants recommended that AI reduce overconfidence and some participants recommended that AI be more transparent.

> *There are many wrong ways and there's a few good ways to implement [AI], and the good ways all involve: Here's how it works, here's what it's looking at, breaking it down, and putting a lot more transparency behind it. –P11*

Participants also suggested methods to build trust in AI-based tools before adoption: allowing practitioners to interactively test with real data, establishing benchmarks for AI-based tools in ICS, adding explanations to predictions, or ensuring that tools were only used as an assistant to human operators.

Around alarms, I would say, starting as an assistant, because there's no way it's going to have all of the experience and all of the information necessary to be 100%. But I think it could do a really good job of helping you. –P7

6 Analysis and recommendations

In this section, we answer our research questions and provide recommendations for adopting AI to support alarm workflows in ICS. We discuss how AI could use existing data and systems for alarms (**RQ1**, Sec. 6.1), how AI could support humans in alarm workflow tasks (**RQ2**, Sec. 6.2), and how to navigate barriers that hinder AI adoption (**RQ3**, Sec. 6.3).

6.1 Deploying AI in systems for alarms

What data and systems are used for alarms in ICS, and are they suitable for AI? (RQ1) We found that the systems and practices for alarms vary across ICS (Sec. 4.1). Alarms operate on process data, network data, or data from cybersecurity tools; alarm logic is programmed into sensors, PLCs, SCADA, or DCS; and alarms can be forwarded to and displayed on various devices. Systems for alarms also vary in the degree of vendor involvement. These differences suggest that, although most prior work that proposes AI-based ICS anomaly detection assumes that all process-level data and compute are available for inference [23, 33], deploying a centralized AI model with real-time access to all process features is unlikely to be feasible for most ICS.

We found that organizations that work with ICS often centralize and store historical data, which may make it suitable for AI. With historical data, participants reported performing data analysis tasks, labeling data, and using automated tools that suggest a readiness for adopting AI (Sec. 4.2).

Recommendation: Consider the varying availabilities of data and infrastructure in ICS when deploying AI. Effectively using AI-based tools to support alarm workflows in ICS requires considering how an AI-based tool would be deployed: what data will be used for training and inference, where inference will be performed, and whether tools will be managed by plant owners or vendors. Designers of AI-based tools need to consider that it is likely that they would be training AI models on only a subset of process-level features.

In domains other than ICS, some deployment models of AI may already fit with the existing systems and practices of some ICS, such as decentralized algorithms for model training and inference (e.g., federated learning across IoT devices [45]) or accessing AI-based tools through a vendor (e.g., AI as a service [50]). Future work should investigate and develop deployment models for AI that match the varying requirements of ICS environments.

Recommendation: Acknowledge the impact of government regulation on AI adoption. We found that differences in government regulation impact alarm workflow practices (Sec. 4.4). For example, in the electric industry, NERC CIP regulations impose security and documentation requirements on connected devices and collected data, which causes some plant owners to choose not to connect certain devices to networks. Developers of AI-based tools for the electric industry will similarly need to comply with NERC CIP regulations. Thus, adopting AI for ICS in industries with stricter regulations (e.g., electricity) will be more difficult than in others (e.g., oil and gas). Researchers and designers of AI-based tools may find new opportunities by focusing on deployment in industries with more flexibility for AI adoption.

6.2 Using AI to support alarm workflow tasks

What human tasks are performed for alarms in ICS, and can AI support them? (RQ2) We found that several tasks are performed for alarms in ICS (Sec. 4.2), as shown in Table 2. In particular, beyond real-time anomaly detection, humans analyze alarms post-hoc for alarm diagnosis and alarm management. Most prior work in AI for ICS security focuses on AI-based anomaly detection [23, 26], but participants reported challenges with alarm diagnosis and alarm management, which rely heavily on intuition and expertise (Sec. 4.3). Participants themselves also suggested that AI could support these tasks (Sec. 5.2). We therefore propose designing an AI-based tool to support alarm diagnosis or alarm management. Given an alarm (or set of alarms), an AI-based tool could help humans triage alarms, suggest potential remediation actions, or predict root causes; or given a larger (e.g., from the past month) dataset of alarm and response data, an AI-based tool could suggest alarms to be added, removed, or modified.

Recommendation: Design AI-based tools to assist humans, rather than act autonomously. As described in Sec. 5.3, we found that participants prefer that AI-based tools make suggestions rather than automate decisions. Furthermore, as described in Sec. 4.3, alarm diagnosis and alarm management are often performed post-hoc and are used to address rare and complex situations; a human-facing, AI-based assistant may be appropriate for these tasks where urgent action is not required. Researchers and designers of AI-based tools should therefore focus on interactively assisting humans. Prior work has explored methods to provide assistance through human-AI interaction [43, 68], and a promising area of future work would be to apply such methods to ICS alarm workflow tasks.

Recommendation: Design AI-based tools to produce unintrusive, actionable outputs. We found that nuisance alarms and operator fatigue often hinder alarm response, and diagnosing unclear and unactionable alarms is a common challenge (Sec. 4.3). We recommend that AI-based tool designers ensure

Stage of Alarm Workflow	Available Input Data	Prediction Task(s)	Anticipated End-user
Anomaly detection	Real-time process data	Detect anomalies in real time	Operator
Alarm diagnosis	Alarm with context	Suggest real-time alarm response Predict root cause of alarm	Operator Lead operator
Alarm management	Set of prior alarms and actions	Fix a misconfigured alarm Improve alarm ruleset	Engineer Manager

Table 2: We suggest opportunities for AI-based tools to support different alarm workflow tasks. For different stages of an alarm workflow (shown in Fig. 3), we list the expected input data available for AI, potential prediction tasks for AI, and anticipated end-users who would use these predictions.

that AI outputs are actionable (e.g., by using AI-based explanations to suggest actions with each prediction [1, 24, 46]), and we recommend that AI-based tools balance the desire to inform the user with giving the user the ability to avoid repeated notifications if they are judged to be incorrect (e.g., by allowing humans to make decisions without AI involvement [8, 13]).

6.3 Navigating barriers to AI adoption

What ICS-specific factors hinder AI adoption? (RQ3) We found that barriers from technology (e.g., limited device connectivity), personnel (e.g., insufficient training), and culture (e.g., tensions between IT and OT professionals) can limit alarm workflow design and execution in ICS (Sec. 4.4). Participants also reported reservations about adopting AI in ICS (Sec. 5.3), stemming from a general mistrust of new technology in ICS (Sec. 4.5). Vendors are said to fail at meeting the requirements of ICS by imposing high costs, not providing adequate customization, or not matching the culture of safety in ICS. To avoid making similar mistakes when deploying AI-based tools to protect ICS, we recommend ways for tool designers to navigate these barriers.

Recommendation: Design AI-based tools for the skill sets of practitioners who work with ICS. We found that alarm workflows involve multiple people with different tasks and specialties (Sec. 4.2). For example, alarm diagnosis could be performed by an operator who first sees the alarm, a manager determining the root cause of an alarm that was incorrectly responded to, or an engineer debugging the logic of a misconfigured alarm. AI-based tools should be designed for specific users performing these tasks, rather than generally for practitioners. Table 2 lists potential opportunities for AI to support different alarm workflow tasks.

Participants also reported that requirements on personnel would be a barrier to adopting AI-based tools (Sec. 5.3), and plant owners are unlikely to hire or train skilled end-users to use AI-based tools. Instead, AI-based tools should be tailored to the existing skill sets of practitioners working with ICS. Prior work in other domains has investigated how explanations of AI outputs can be modified based on levels of user expertise [14, 20], and a promising area of future work would be to apply such approaches to support practitioners in ICS.

Recommendation: To build trust in AI, focus on demonstrating transparency to users. We found that trust and reputation were more important than quantitative metrics when deciding to adopt new technology in ICS (Sec. 4.5). Practitioners who work with ICS often do not currently trust AI (Sec. 5.3), so AI-based tool designers must first build this trust. Participants suggested building trust by transparently demonstrating how AI-based tools make decisions. Furthermore, participants reported that OT professionals face challenges working with IT professionals, are weary of new technology being pushed into their environments, and can be excluded from tool-adoption decisions (Sec. 4.4). Thus, AI-based tool demonstrations should include practitioners who work with ICS and OT.

As a first step to build trust in AI, we recommend pilot projects that allow interactive testing of AI-based tools and focus on transparency of AI models. Prior work has developed interactive tools for prototyping AI-based tools by allowing users to modify data and observe changes in predictions [64, 65], and future work should investigate whether such methods are effective for practitioners who work with ICS.

7 Conclusion

We investigated current practices for alarms in ICS and identified challenges and opportunities for AI to support these practices. After conducting semi-structured interviews with 18 practitioners who work on safeguarding and securing ICS in different roles, from performing alarm response to building tools for alarms, we identified opportunities to adopt AI-based tools to support alarm diagnosis and alarm management. Finally, we recommend ways for researchers and designers of AI-based tools to navigate barriers to adoption in ICS, such as considering AI models with access to only a subset of process-level features and interactively demonstrating model transparency to practitioners.

Acknowledgments

We thank the anonymous reviewers for their feedback in improving this work; Brian Singer and Andy Gallardo for their help in developing our interview script; and Eunsuk Kang and Amritanshu Pandey for their help in participant recruitment. This paper is based on work supported in part by Mitsubishi Heavy Industries through the Carnegie Mellon CyLab partnership program.

References

[1] Sridhar Adepu, Nianyu Li, Eunsuk Kang, and David Garlan. Modeling and analysis of explanation for secure industrial control systems. *ACM Transactions on Autonomous and Adaptive Systems*, 17(3-4), 2022.

[2] Bushra A. Alahmadi, Louise Axon, and Ivan Martinovic. 99% false positives: A qualitative study of SOC analysts' perspectives on security alarms. In *31st USENIX Security Symposium (USENIX Security 22)*, 2022.

[3] Mennatallah Amer, Markus Goldstein, and Slim Abdennadher. Enhancing one-class support vector machines for unsupervised anomaly detection. In *ACM SIGKDD Workshop on Outlier Detection and Description*, 2013.

[4] Wissam Aoudi, Mikel Iturbe, and Magnus Almgren. Truth will out: Departure-based process-level detection of stealthy attacks on control systems. In *ACM SIGSAC Conference on Computer and Communications Security*, 2018.

[5] Mohammed Asiri, Neetesh Saxena, Rigel Gjomemo, and Pete Burnap. Understanding indicators of compromise against cyber-attacks in industrial control systems: A security perspective. *ACM Transactions on Cyber-Physical Systems*, 2023.

[6] Rima Asmar Awad, Saeed Beztchi, Jared M. Smith, Bryan Lyles, and Stacy Prowell. Tools, techniques, and methodologies: A survey of digital forensics for SCADA systems. In *4th Annual Industrial Control System Security Workshop*, 2018.

[7] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The menlo report. *IEEE Security & Privacy*, 10(2), 2012.

[8] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S Lasecki, Daniel S Weld, and Eric Horvitz. Beyond accuracy: The role of mental models in human-AI team performance. In *AAAI conference on Human Computation and Crowdsourcing*, volume 7, 2019.

[9] Margret Bauer, Alexander Horch, Lei Xie, Mohieddine Jelali, and Nina Thornhill. The current state of control loop performance monitoring–a survey of application in industry. *Journal of Process Control*, 38, 2016.

[10] Jack Beerman, David Berent, Zach Falter, and Suman Bhunia. A review of colonial pipeline ransomware attack. In *IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW)*, 2023.

[11] Deval Bhamare, Maede Zolanvari, Aiman Erbad, Raj Jain, Khaled Khan, and Nader Meskin. Cybersecurity for industrial control systems: A survey. *Computers & Security*, 89:101677, 2020.

[12] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 2006.

[13] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *ACM on Human-Computer Interaction*, 5(CSCW1), 2021.

[14] Valerie Chen, Q Vera Liao, Jennifer Wortman Vaughan, and Gagan Bansal. Understanding the role of human intuition on reliance in human-AI decision-making with explanations. *ACM on Human-computer Interaction*, 7(CSCW2), 2023.

[15] Jason D. Christopher. Sans 2024 state of ICS/OT cybersecurity. *SANS Institute*, 2024.

[16] Oakley Cox. Three ways AI secures OT & ICS from cyber attacks, 2024. https://www.darktrace.com/blog/three-ways-ai-secures-operational-technology-ot-industrial-control-systems-ics-from-cyber-attacks.

[17] Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *AAAI Conference on Artificial Intelligence*, volume 35, 2021.

[18] Berend Denkena, M-A Dittrich, Hendrik Noske, and Matthias Witt. Statistical approaches for semi-supervised anomaly detection in machining. *Production Engineering*, 14:385–393, 2020.

[19] Willian Dimitrov and Svetlana Syarova. Analysis of the functionalities of a shared ICS security operations center. In *2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE)*, 2019.

[20] Upol Ehsan, Samir Passi, Q Vera Liao, Larry Chan, I-Hsiang Lee, Michael Muller, and Mark O Riedl. The who in XAI: How AI background shapes perceptions of AI explanations. In *CHI Conference on Human Factors in Computing Systems*, 2024.

[21] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring how privacy and security factor into IoT device purchase behavior. In *CHI Conference on Human Factors in Computing Systems*, 2019.

[22] Cheng Feng, Venkata Reddy Palleti, Aditya Mathur, and Deeph Chana. A systematic framework to generate invariants for anomaly detection in industrial control systems. In *Network and Distributed System Security Symposium*, 2019.

[23] Clement Fung, Shreya Srinarasi, Keane Lucas, Hay Bryan Phee, and Lujo Bauer. Perspectives from a comprehensive evaluation of reconstruction-based anomaly detection in industrial control systems. In *27th European Symposium on Research in Computer Security*, 2022.

[24] Clement Fung, Eric Zeng, and Lujo Bauer. Attributions for ML-based ICS anomaly detection: From theory to practice. In *31st Network and Distributed System Security Symposium*, 2024.

[25] Andrea Gallardo, Robert Erbes, Katya Le Blanc, Lujo Bauer, and Lorrie Faith Cranor. Interdisciplinary approaches to cybervulnerability impact assessment for energy critical infrastructure. In *CHI Conference on Human Factors in Computing Systems*, 2024.

[26] Jairo Giraldo, David Urbina, Alvaro Cardenas, Junia Valente, Mustafa Faisal, Justin Ruths, Nils Ole Tippenhauer, Henrik Sandberg, and Richard Candell. A survey of physics-based attack detection in cyber-physical systems. *ACM Computing Surveys*, 51(4):1–36, 2018.

[27] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. ït's a scavenger hunt": Usability of websites' opt-out and data deletion choices. In *CHI Conference on Human Factors in Computing Systems*, 2020.

[28] Dina Hadžiosmanović, Robin Sommer, Emmanuele Zambon, and Pieter H. Hartel. Through the eye of the PLC: Semantic security monitoring for industrial processes. In *30th Annual Computer Security Applications Conference*, 2014.

[29] Anna Hall and Vivek Agarwal. Barriers to adopting artificial intelligence and machine learning technologies in nuclear power. *Progress in Nuclear Energy*, 175, 2024.

[30] Bill R. Hollifield. Understanding & applying the ANSI-ISA 18-2 alarm management standard. Technical report, Hexagon, 2023.

[31] Albert T. Jones and Charles R. McLean. A proposed hierarchical control model for automated manufacturing systems. *Journal of Manufacturing Systems*, 5(1), 1986.

[32] Taewook Kim, Hyomin Han, Eytan Adar, Matthew Kay, and John Joon Young Chung. Authors' values and attitudes towards AI-bridged scalable personalization of creative language arts. In *CHI Conference on Human Factors in Computing Systems*, 2024.

[33] Abigail MY Koay, Ryan K L Ko, Hinne Hettema, and Kenneth Radke. Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges. *Journal of Intelligent Information Systems*, 60(2), 2023.

[34] Faris Bugra Kokulu, Ananta Soneji, Tiffany Bao, Yan Shoshitaishvili, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. Matched and mismatched SOCs: A qualitative study on security operations center issues. In *ACM SIGSAC Conference on Computer and Communications Security*, 2019.

[35] Moshe Kravchik and Asaf Shabtai. Efficient cyber attack detection in industrial control systems using lightweight neural networks and PCA. *IEEE Transactions on Dependable and Secure Computing*, 19(4), 2022.

[36] Nir Kshetri and Jeffrey Voas. Hacking power grids: A current problem. *Computer*, 50(12), 2017.

[37] Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. In *International Conference on Artificial Neural Networks*, 2019.

[38] Sara Ljungblad, Yemao Man, Mehmet Aydın Baytaş, Mafalda Gamboa, Mohammad Obaid, and Morten Fjeld. What matters in professional drone pilots' practice? an interview study to understand the complexity of their work and inform human-drone interaction research. In *CHI Conference on Human Factors in Computing Systems*, 2021.

[39] Yuan Luo, Ya Xiao, Long Cheng, Guojun Peng, and Danfeng Yao. Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities. *ACM Computing Surveys*, 54(5):1–36, 2021.

[40] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *ACM on Human-Computer Interaction*, 3(CSCW), 2019.

[41] Stephen McLaughlin, Charalambos Konstantinou, Xueyang Wang, Lucas Davi, Ahmad-Reza Sadeghi, Michail Maniatakos, and Ramesh Karri. The cybersecurity landscape in industrial control systems. *Proceedings of the IEEE*, 104, 2016.

[42] Jaron Mink, Hadjer Benkraouda, Limin Yang, Arridhana Ciptadi, Ali Ahmadzadeh, Daniel Votipka, and Gang Wang. Everybody's got ML, tell me what else you have: Practitioners' perception of ML-based security tools and explanations. In *IEEE Symposium on Security and Privacy*, 2023.

[43] Hussein Mozannar, Gagan Bansal, Adam Fourney, and Eric Horvitz. When to show a suggestion? integrating human feedback in AI-assisted programming. In *AAAI Conference on Artificial Intelligence*, volume 38, 2024.

[44] Glenn Murray, Michael N Johnstone, and Craig Valli. The convergence of it and ot in critical infrastructure. In *15th Australian Information Security Management Conference*, 2017.

[45] Dinh C Nguyen, Ming Ding, Pubudu N Pathirana, Aruna Seneviratne, Jun Li, and H Vincent Poor. Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys and Tutorials*, 23(3), 2021.

[46] Megan Nyre-Yu, Elizabeth Morris, Michael Smith, Blake Moss, and Charles Smutz. Explainable ai in cybersecurity operations: Lessons learned from XAI tool deployment. In *Usable Security and Privacy (USEC) Symposium*, 2022.

[47] Cliodhna O'Connor and Helene Joffe. Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 2020.

[48] Sean Oesch, Robert Bridges, Jared Smith, Justin Beaver, John Goodall, Kelly Huffer, Craig Miles, and Dan Scofield. An assessment of the usability of machine learning based tools for the security operations center. In *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*, 2020.

[49] Dean Parsons. SANS ICS/OT cybersecurity survey: 2023's challenges and tomorrow's defenses. *SANS Institute*, 2023.

[50] Mauro Ribeiro, Katarina Grolinger, and Miriam AM Capretz. MLaaS: Machine learning as a service. In *IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015.

[51] Robert M. Lee, Michael J. Assante and Tim Conway. Analysis of the cyber attack on the Ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 388, 2016.

[52] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52:1893–1907, 2018.

[53] Wenli Shang, Peng Zeng, Ming Wan, Lin Li, and Panfeng An. Intrusion detection algorithm based on OCSVM in industrial control system. *Security and Communication Networks*, 9(10), 2016.

[54] Alex Shultz. For the first time, artificial intelligence is being used at a nuclear power plant: California's diablo canyon. The Markup, 2025. https://themarkup.org/artificial-intelligence/2025/04/08/for-the-first-time-artificial-intelligence-is-being-used-at-a-nuclear-power-plant-californias-diablo-canyon.

[55] Brian Singer, Amritanshu Pandey, Shimiao Li, Lujo Bauer, Craig Miller, Lawrence Pileggi, and Vyas Sekar. Shedding light on inconsistencies in grid cybersecurity: Disconnects and recommendations. In *IEEE Symposium on Security and Privacy*, 2023.

[56] Pooja Singh and Lalit Kumar Singh. Instrumentation and control systems design for nuclear power plant: An interview study with industry practitioners. *Nuclear Engineering and Technology*, 53(11), 2021.

[57] Joseph Slowik. Evolution of ICS attacks and the prospects for future disruptive events. *Threat Intelligence Centre Dragos Inc*, 2019.

[58] Lei Song, Chuheng Zhang, Li Zhao, and Jiang Bian. Pre-trained large language models for industrial control. *arXiv preprint arXiv:2308.03028*, 2023.

[59] Kirti Soni, Nishant Kumar, Anjali S Nair, Parag Chourey, Nirbhow Jap Singh, and Ravinder Agarwal. Artificial intelligence: Implementation and obstacles in industry 4.0. In *Handbook of Metrology and Applications*. Springer, 2022.

[60] Keith Stouffer. Guide to industrial control systems (ICS) security. *NIST special publication*, 800(82), 2011.

[61] Eric Stranz and Stefan Nohe. NERC CIP in the real world on a real budget. In *Minnesota Power Systems Conference*, 2016.

[62] David I. Urbina, Jairo A. Giraldo, Alvaro A. Cardenas, Nils Ole Tippenhauer, Junia Valente, Mustafa Faisal, Justin Ruths, Richard Candell, and Henrik Sandberg. Limiting the impact of stealthy attacks on industrial control systems. In *ACM SIGSAC Conference on Computer and Communications Security*, 2016.

[63] Mathew Vermeer, Natalia Kadenko, Michel van Eeten, Carlos Gañán, and Simon Parkin. Alert alchemy: SOC workflows and decisions in the management of NIDS rules. In *ACM SIGSAC Conference on Computer and Communications Security*, 2023.

[64] Zijie J Wang, Alex Kale, Harsha Nori, Peter Stella, Mark E Nunnally, Duen Horng Chau, Mihaela Vorvoreanu, Jennifer Wortman Vaughan, and Rich Caruana. Interpretability, then what? editing machine learning models to reflect human knowledge and values. In *28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022.

[65] Zijie J Wang, Jennifer Wortman Vaughan, Rich Caruana, and Duen Horng Chau. GAM coach: Towards interactive and user-centered algorithmic recourse. In *CHI Conference on Human Factors in Computing Systems*, 2023.

[66] Joseph Weiss, Rob Stephens, and Nadine Miller. Changing the paradigm of control system cybersecurity. *Computer*, 55, 2022.

[67] Joseph Weiss, Rob Stephens, and Nadine Miller. Control system cyber incidents are real—and current prevention and mitigation strategies are not working. *Computer*, 55, 2022.

[68] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *CHI Conference on Human Factors in Computing Systems*, 2020.

[69] Alberto Zanutto, Benjamin Oliver Shreeve, Karolina Follis, Jeremy Simon Busby, and Awais Rashid. The shadow warriors: In the no man's land between industrial control systems and enterprise IT systems. In *Symposium on Usable Privacy and Security*, 2017.

[70] Kim Zetter. A cyberattack has caused confirmed physical damage for the second time ever. WIRED, 2015. https://www.wired.com/2015/01/german-steel-mill-hack-destruction/.

A Semi-structured interview script

The detailed questions used in our semi-structured interview script are provided below. When appropriate, we asked follow up questions to participants to encourage further elaboration.

Part I - Background information
Participant demographics:

- What type(s) of ICS do you work on?
- What is your job title?
- How many years of experience do you have with ICS/OT?
- Do you have a background in cybersecurity? Describe/how many years?
- Do you have a background in machine learning? Describe/how many years?

Specifics of ICS:

- What parts of the system are monitored and controlled by PLCs, SCADAs, etc?
- What types of data is collected, how is it collected, and how is it shown to a human?
- Could you describe your day-to-day responsibilities?

Part II - Questions about monitoring and alarms at your organization
Addressing issues in ICS:

- What are examples of anomalies or problems in the process that would raise alarms in your ICS?
- How are these issues detected?
- When an alarm is raised, what is your role in addressing the situation?
- How do you find out about these situations? (Are you watching an HMI? Do you get assigned work orders?)
- Who around you is involved in addressing these situations? (Are you managing people who address it? Are other people analyzing the situation and asking you to investigate?)

Anomaly detection tools:

- What tools or systems does your organization use to detect anomalies?
- Does your system primarily rely on rules, ML, or both?
- Does your operate on network information (e.g., packets), operations information (e.g., sensors and actuators), or both?

Responsibility around anomaly detection:

- Who sets up anomaly detectors, rules, or set points? You, other people in your org, outside vendors?
- What types of data sources are used in the monitoring system?
- Who is responsible for these sources? How many people in the org are in that role?
- What types of actions are commonly taken or expected, based on information from the monitoring system?
- Who is responsible for taking these actions? How many people in the org are in that role?

Responding to anomalies:

- How many alarms do you receive, and what proportion of them do you have to manually investigate?
- Do alarms need to be triaged? Is this process difficult?
- What starts an investigation of an alarm?
- How do you determine if the alarm in question is a false positive? Walk through your process in making this decision.
- Are any tools used to help with diagnosis, what information is provided by this tool?
- How is information from this tool used to determine if an alarm is a false positive?
- How is information from this tool be used to determine next steps for remediation?
- What other information would you need about an alarm to help determine if it is a false positive or not?
- Could parts of this decision be automated? Would you trust it?

General perspectives:

- What are the most helpful or useful aspects of the tools you use?
- What are the main challenges in detecting and debugging anomalous behavior?
- What are the main challenges in working with ICS in general?

Part III - Tool adoption

- How did your organization decide on the tools it uses for monitoring?
- Are there other roles in the organization who measure these things, experiment with tools, or deploy them?
- What properties or metrics were used to distinguish the tools you use from other alternatives?
- Are these decisions based on quantitative metrics? If no, then what is it based on?
- Once a tool is deployed, do you use any metrics or processes to ensure that it is useful for your organization?
- Suppose a vendor suggests that you try a new tool, what properties would it need to have for you to consider adopting it?

Part IV - Perceptions of AI

- What do you consider to be the pros and cons of using AI-based anomaly detection methods vs traditional methods for alarms in ICS?
- What improvements would need to be made to AI to make it more trusted by your organization?

Part V - Miscellaneous

- Is there anything else you wanted to tell us that we didn't ask about?

B Codebook

We provide our codebook in Table 3 and Table 4. The codes shown in Table 3 are used to analyze the responses in part II of our interview script, and the codes shown in Table 4 are used to analyze the responses in part III and part IV of our interview script; additional codes are included in Table 4 for general themes that emerged, which were not specific to an interview question.

Name of Code	Description	# Matched
role > OTcybersecurity	Performs OT cybersecurity tasks in their role	5
role > alarmResponse	Performs alarm response tasks in their role	4
role > engineering	Performs engineering tasks in their role	11
role > manager	Performs management tasks in their role	6
role > operations	Performs operations tasks in their role	4
teamsize	Details about team size	6
architecture > usesPLC	Organization uses PLCs	12
architecture > usesDCS	Organization uses a DCS	5
architecture > usesSCADA	Organization uses SCADA	5
architecture > usesHMI	Organization uses HMIs	7
architecture > usesMainControlRoom	Organization uses a control room	6
architecture > usesSubControlRoom	Organization uses multiple control rooms	2
architecture > usesHistorian	Organization uses a data historian	7
architecture > detailsPLC	Details about how PLCs are used	10
architecture > detailsSCADA	Details about how SCADA/DCS are used	14
alarmArch > PLCs	Alarms come from PLCs	13
alarmArch > SCADA	Alarms come from SCADA/DCS	6
alarmArch > safetySystem	Alarms come from a safety system	11
alarmArch > external	Alarms come from an external tool	2
alarmDefn > bounds	Alarms defined as upper/lower bounds	12
alarmDefn > custom	Alarms defined as custom logic	10
alarmRole > operations	Operators configure alarms	3
alarmRole > team	A team configures alarms	3
alarmRole > vendor	A vendor configures alarms	4
alarmTypes > process	Alarms for unwanted process values	16
alarmTypes > communication	Alarms for communication issues	6
alarmTypes > componentFailure	Alarms for component failures	6
alarmTypes > cybersecurity	Alarms for cybersecurity issues	3
alarmTypes > physicalSecurity	Alarms for physical security issues	3
alarmTypes > other	Alarms for other types of issues	5
alarmResponse > triage	Details about triage process in alarm response	13
alarmResponse > controlRoom	Details about control rooms in alarm response	4
alarmResponse > severity	Details about severity levels in alarm response	9
alarmResponse > operations	Details about coordination with operators in alarm response	11
alarmResponse > humanFactors	Details about human factors in alarm response	6
alarmResponse > UI	Details about user interfaces in alarm response	5
alarmNumber	Details about number/rate of alarms	9
alarmPhilosophy	Details about what should be an alarm	9
alarmActionability	Details about actionability of alarms	4
alarmDiagnosis > challenges	Details about challenges when diagnosing alarms	12
alarmDiagnosis > nuisance	Details about nuisance alarms when diagnosing alarms	12
alarmDiagnosis > intuition	Details about need for intuition when diagnosing alarms	11
alarmDiagnosis > postHoc	Details about post hoc analysis of alarms	10
alarmDiagnosis > tools	Details about vendor tools when diagnosing alarms	8
alarmDiagnosis > ML	Details about using ML to diagnose alarms	3
alarmManage > meeting	Details about alarm management meeting	4
alarmManage > testing	Details about testing alarms	5
alarmManage > update	Details about updating alarms	5

Table 3: Codes for responses in Part II of our interview script, which focuses on tasks and systems in alarm workflows. For each code, we provide: its name and structure, its description, and the number of participants matched to it.

Name of Code	Description	# Matched
tools > barriersCultural	Barriers to tool adoption from ICS culture	10
tools > barriersTechnical	Barriers to tool adoption from ICS technical limitations	6
tools > peopleExcluded	Details about people excluded in adoption	4
tools > peopleIncluded	Details about people included in adoption	5
tools > values	Details about important values for adoption	16
tools > metrics	Details about how metrics are used to evaluate tools	13
AI > positive	General positive perceptions of AI	15
AI > positive > saveTime	AI will save time	9
AI > positive > complex	AI performs complex tasks better than humans	9
AI > positive > exciting	AI is novel and exciting	3
AI > negative	General negative perceptions of adopting AI	17
AI > negative > trust	ICS would not trust AI	11
AI > negative > criticality	ICS are too critical for AI	10
AI > negative > complex	Parts of ICS are too complex for AI	5
AI > negative > transparency	AI decisions are not transparent	16
AI > negative > dataCost	AI requires data that we do not have	8
AI > negative > moneyCost	AI requires money that we do not have	4
AI > negative > peopleCost	AI requires people than we do not have	6
AI > negative > badAI	Negative perceptions of AI itself	5
AI > conceptual	Conceptual models of AI	18
AI > conceptual > LLM	Talked about LLMs	8
AI > conceptual > neuralNet	Talked about neural networks	5
AI > conceptual > linear	Talked about linear regression or classification	3
AI > conceptual > prediction	Talked about data for AI predictions generally	6
AI > conceptual > training	Talked about data for AI training generally	12
AI > useCase > assistant	Suggest to use AI as an assistant	7
AI > useCase > optimizeProcess	Suggest to use AI to optimize process	8
AI > useCase > optimizeAlarms	Suggest to use AI to optimize alarm workflows	8
AI > useCase > maintenance	Suggest to use AI for system maintenance	3
AI > recommendations	Recommendations for how AI should improve	8
external > alarms	An external party performs part of alarm workflow actions	1
external > implementation	An external party implements part of the alarm workflow	4
external > detailsExternal	Details about contracts with external parties	9
compareIndustry > cultural	Comparing ICS industries based on cultural differences	5
compareIndustry > longitudinal	Comparing ICS based on trends over time	6
compareIndustry > size	Comparing ICS based on their size	5
compareIndustry > technical	Comparing ICS industries based on technical differences	4
misc > cybersecurityPerceptions	Current cybersecurity perceptions in ICS	9
misc > cybersecurityPractices	Current cybersecurity practices in ICS	9
misc > regulations	Government regulations affecting ICS	7
misc > cultureClash	IT/OT culture clash	8
misc > painPeople	Personnel issues affecting ICS	8
misc > painTechnical	Technical issues affecting ICS	9

Table 4: Codes for responses in Part III and Part IV of our interview scripts, which focus on vendor tool adoption (Part III) and perceptions of AI (Part IV). We also include codes for other miscellaneous themes, such as cross-industry and cross-functional pain points. For each code, we provide: its name and structure, its description, and the number of participants matched to it.